

UNITED STATES OF AMERICA

v.

Manning, Bradley E.
PFC, U.S. Army,
HHC, U.S. Army Garrison,
Joint Base Myer-Henderson Hall
Fort Myer, Virginia 22211

STIPULATION OF
EXPECTED TESTIMONY

Mr. James Downey

17 June 2013

It is hereby agreed by the Accused, Defense Counsel, and Trial Counsel, that if Mr. James Downey were present to testify during the merits and pre-sentencing phases of this court-martial, he would testify substantially as follows.

1. I work at Defense Information Systems Agency (DISA), Fort Meade, Maryland. Specifically, I am a part of the Program Executive Office for Mission Assurance (PEO-MA) and Network Operations. I am the program manager for attack analysis. I have held this position since 2007. I hold the Global Information Assurance Certification (GIAC) security leadership certification (GSLC), and I am a certified ethical hacker.

2. The PEO-MA department, where I currently work, provides program management for various programs that help secure the IT information within the Department of Defense (DoD). Within PEO-MA, I work for the Community Data Center (CDC). The CDC hosts a set of tools used by people who secure DoD networks. We host enterprise level Information Assurance (IA) tools and net defense tools, which enable analysts to basically ensure the availability and integrity of the networks that DISA provides for DoD. "Enterprise tools" are those which are capable of handling the amount of data we deal with and the large and complicated networks with which we work. Since DISA is like an internet server provider for DoD, we operate on a scale which is much larger than what the commercially available tools are designed to handle. A "tool" is just what it sounds like – something that allows us to do our network management job. Usually, it is information or a way of processing or gathering information.

3. The tool relevant to this case is the data we use called Netflow data. This is a type of data which was developed by Cisco, but which is now industry standard. With it, we can capture the Internet Protocol (IP) addresses of two computers communicating across the system, as well as the volume of traffic which flows between them. We use YAF to collect this data. YAF stands for "yet another flow meter". This tool was developed by Carnegie Mellon and is the industry standard. Just like any meter, it measures and then creates a data record of the flow past a data collection point. A point of collection is any of the various monitoring points we have stationed at key perimeter locations throughout the network; for example, where a DoD network crosses or connects to the commercial world. These points monitor all traffic, or computer to computer communication, crossing from one side of the router to another. Our system would "see" when someone is on a work computer browsing internet websites like yahoo while on NIPRNET or the United States Central Command Server from a computer in Iraq while on SIPRNET. The system detects the capacity being used during that communication and at that location. This information can communicate whether something is being downloaded onto that computer. We collect

Netflow data on NIPRNET and SIPRNET. There are relatively few routers collecting Netflow data throughout the entire SIPRNET. Because this system only collects information passing from one side of the router to another, it does not collect Netflow data passing within a network that does not cross through a collecting router. For example, if a computer is communicating with a server or another computer within the Iraq SIPRNET domain, that activity would not be captured in Netflow data, because the connection and data does not cross through the Iraq SIPRNET domain router, but rather stays within the Iraq SIPRNET domain network.

4. We collect this Netflow data for several reasons. First, we use the data to conduct traffic analysis. It allows our analysts to see where they need to deploy additional capacity in the DoD network. For example, if one segment of the system is getting more traffic than another, it may need a larger router. This type of work falls within the purview of those CDC analysts working on network operations and maintenance. These analysts focus on maintaining the availability, robustness, and proper functioning of the Netflow data. They ensure that the system is collecting data correctly, that it is securely transported and stored, and that the system used to access the information is on-line and functioning properly. Another section however uses the Netflow data to defend the DoD network from threats. For example, if a regular DoD user has a virus on his/her computer that tries to connect to a malicious computer outside the network, our tools enable our analysts to detect that and take the offending computer offline. This section can investigate suspicious activity. I work in the section that manages the delivery of CDC capabilities. Finally, we also have a group using the Netflow data to do research and development. The Research and Development group analyzes the data to try and find patterns which might help them identify behavior going on that we do not currently have a means of detecting. For example, with older viruses it's easy to know when something has been infected, but newer ones can be more cautious in how they operate. By looking at patterns over time, our analysts might be able to see something that helps them find compromised computers in the network before the virus infects others.

5. CENTAUR is what we call the system we use to track the Netflow data I just described. It is one of the systems with which I work in the PEO-MA. A CENTAUR log is the data output from our Netflow data system. I became involved in this case after DISA launched an audit initiative focused on integrating and analyzing multiple data sources to identify and track potential insider threats on SIPRNET. Because of my job, I am familiar with the Netflow data CENTAUR generates for SIPRNET. I received a request to pull the CENTAUR logs showing communication between three SIPRNET IPs and for a specific period of time. The date range of interest was October 2009 to May 2010. Investigators were interested in the following IP addresses: 22.225.29.185, 22.225.41.22, and 22.225.41.40.

6. To look at the CENTAUR data from SIPRNET, analysts use a tool developed by Carnegie Melon, called SILK – System for Internet Level Knowledge. Once gathered, the CENTAUR log can show certain pieces of information. I will explain how to read the CENTAUR logs by way of example with the following log:

| sIP | dIP | sPort | dPort | pro | packets | bytes | flags | sTime | dur | eTime | sensor |
|---------------|--------------|-------|-------|-----|---------|---------|----------|-------------------------|---------|-------------------------|--------------|
| 204.37.126.39 | 22.225.41.40 | 80 | 2641 | 6 | 1379 | 1305267 | FS PA | 2009/12/19T01:41:43.633 | 112.650 | 2009/12/19T01:43:36.283 | SPE- SMEC |

a. The "sIP" is the source IP. It is the Internet Protocol (IP) address of the computer that initiated the conversation that log line is tracking. A "conversation" is a set of transactions that has in common the same source and destination IPs and ports and which occurs within the same time frame. In the above example, 22.225.41.40 (the destination IP addressed) received information from the sending IP address, which is 204.37.126.39. **Prosecution Exhibit (PE) 152** ^{AF} _{164 Rem} 111 **for Identification** is a list of many of the organizations who are associated with or own the IP addresses searched by the 22.225.29.185, 22.225.41.22, and 22.225.41.40 IP addresses.

b. The "dIP" or destination IP is the IP address for the computer that received the data from the sIP. The dIP in the above example is 22.225.41.40.

c. The "sPort" is the port that the sIP was using to communicate. A port itself is a way the computer can carry on multiple conversations on a network at the same time. You can think of it like a mail slot or a particular channel that a computer uses to hold a conversation. The sPort in the above example is 80. sPort code 80 indicates all internet web traffic, including browsing on the web.

d. The "dPort" is the destination port. This is the port the dIP was using to communicate. Essentially it is the computer which received the conversation. The dPort in the above example is 2641.

e. The log item "pro" stands for protocol. The protocol is the convention, or language, which the two computers were using to talk to one another. The number "6" is the Transmission Control Protocol (TCP). TCP is a language. TCP is the dominant protocol. Knowing the protocol is important because it tells you the kind of conversation the two logged IP addresses were having. For example, another protocol is the number "1" for ICMP. Protocols like TCP are generally used by users to generate and receive data. Protocols like ICMP are used by computer systems to report back on status or to support other protocols.

f. "Packets" are the chunks a computer breaks information up into in order to transmit it across the network. The ratio of packets to bytes for example can tell analysts about the nature of the conversation occurring; essentially, packets communicate complexity. A byte is simply a unit of measuring the size of data or seeing volume. A large number of bytes relative to the packets means a large file is getting downloaded. A small byte count means a lower level form of communication. The packets in the above example is 1379.

g. A byte itself is simply a unit of measuring the size of data or seeing volume. Beyond its relationship to a packet, in raw form, the byte tells exactly how much information was exchanged in a given conversation. "Bytes" provide a straight forward measurement of how much data was transmitted, where the packets can tell you how that data was transmitted. The bytes in the above example is 1305267.

h. "sTime" is the time the communication between IPs started. In the above example the particular communication between 204.37.126.39 and 22.225.41.40 started at 2009/12/19T01:41:43.633.

i. The column labeled "dur" is its duration. The duration is given in seconds. Knowing this information is important because different durations are characteristic of different kinds of conversations. This information helps analysts like me guide our inferences about the data by providing context for the communication we are seeking to understand. The above communication took 112.650 seconds to complete.

j. "eTime" is the time a communication ends. The above communication between 204.37.126.39 and 22.225.41.40 ended at 2009/12/19T01:43:36.283, which should be 112.650 seconds after the sTime.

k. "Sensor" means the collection point used to collect the data being communicated in the log line. It identifies the router or the sensor generating the record and basically tells us where on the network the traffic the log line describes occurred. The sensor in the above example is SPE-SMEC.

l. Together, this information allows analysts to see the IP address of an individual computer using the system and the complexity and volume of information being communicated as well as the length of time the computer is conducting its activity. The above log tells me that the IP address 22.225.41.40 received 1305267 bytes of data from IP address 204.37.126.39 on 19 December 2009.

7. As I indicated earlier, the CENTAUR log data is very useful in detecting suspicious activity. While it will not automatically alert analysts in the security section of suspicious activity, part of their job is to schedule scripts which look for activity. DISA also deploys detection tools at multiple locations through the network with unique signatures. These look for a particular type of suspicious user activity. For example, going to known blocked websites or known malware servers is something we can detect. When a user takes the action that fits the signature activity, this action triggers an alert to an analyst in something close to real time. Security analysts also work off of tips. These tips can come from digital alerts like the ones I just described or from sources such as Information Assurance groups within DoD organizations which are responsible for their own local security. Alternatively, law enforcement can request information from our system – as occurred in this case.

8. The format in which the system returns results to our queries varies depending on the query. All of these tools run from a command line using text command. The answer to the query comes back in a native binary file. But then another tool translates that automatically into a regular text file so that it is readable by a human. The analyst then defines which information they want to see and in what order. By "information," I am referring to the terms I defined earlier, such as "sIP" and "dur." An analyst then makes the report using this data. The report includes the data and may also include the analyst's interpretation of what that data means, why it's important, and what the context is.


9. The latter occurred in this case. When asked for a certain date range of data tied to the relevant IP addresses in this case, we pulled the data. We found communications for the IP addresses I discussed earlier between November of 2009 and May of 2010. I am not aware of any irregularities occurring, and we did some tests to ensure the data was accurate.


10. Drastic changes in the history of a log tell me one of two things. When a log is not consistent with previous behavior over a large period of time, it would indicate to me that either a sensor was down or the relevant computer was turned completely off. There should always be some baseline level of activity for a computer connected to a network.

11. After collecting the data, I saved files of the log data which were then burned to a CD for the investigators. The CD was marked "6/15/2012, UNCLASSIFIED, hub_out_dip.csv; hub_out_sip.csv; spe_out_dip.csv; spe_out_sip.csv." These .csv file names represent the different log data that was pulled. They show activity of the 22.225.29.185, 22.225.41.40, and 22.225.41.22 IP addresses as the source and destination IPs. The information was sent via tracked FedEx package to Special Agent David Shaver. The tracking number was 875027891920. **PE 152 for ID** is the CD containing the log data I collected. I recognize the data on the CD because I collected it, and I recognize the logs based on the column identifiers and familiarity with CENTAUR logs, which I described earlier. A records custodian attested to their authenticity on 15 June 2012 at BATES number: 00449443.

12. At no point during my collection or transport of these logs did I alter them in any way. I have no reason to believe any of my colleagues altered the data or experienced anything out of the ordinary in collecting it. And, I have no reason to believe the data provided or the device on which it was stored was damaged or contaminated in any way. Finally, I am not aware of any issue in the collection, storage, or transport of this information which would cause it to have been incorrectly preserved.


ANGEL M. OVERGAARD
CPT, JA
Assistant Trial Counsel


THOMAS F. HURLEY
MAJ, JA
Military Defense Counsel


BRADLEY E. MANNING
PFC, USA
Accused